The Impact of Research and Development Expenditure on Unemployment Rate

Georgia Institute of Technology
ECON 3161 Econometrics

Adit Jindal

Abstract
Countries often commit to increasing their expenditure on Research and Development (R&D) as a way to boost economic growth and counteract increasing unemployment rates. There exist few independent studies showing the impact of R&D expenditure on economic growth, and even fewer showing impact on the unemployment rate. This study attempts to uncover the relationship between the unemployment rate as a percentage of the labor force (UE) and the R&D expenditure as a percentage of gross domestic product (GERD) by analyzing data from 71 countries. The other explanatory variables that are taken into consideration are the education level (UNDP Education Index), inflation, economic growth (percentage GDP growth), total population, compensation of employees, manufacturing sector's value added (% of GDP), and service sector's value added (% of GDP). Further, a time delay of three years has been purposefully added to allow the impact of R&D expenditure to manifest itself and impact the unemployment rate. The R&D data is collected from 2016 and the unemployment rate is collected from 2019. The latest data has not been considered due to the overarching effects of the COVID-19 pandemic. Using both single and multiple linear regression models, a negative correlation was found between the unemployment rate and the R&D expenditure of a country.

I. Introduction

Research and Development (R&D) refers to the activities done to innovate and introduce new products and services in the economy. According to the United Nations Sustainable Development Goal 9 target 5, countries have pledged to increase the number of research and development workers per 1 million people, and public and private research by 2030. Hence, global spending on R&D has surpassed almost US $1.7 trillion.

This extensive expenditure is being done assuming that more R&D research would boost the economy, and reduce the unemployment rate. The COVID-19 pandemic created an unprecedented economic crisis with the unemployment rate surging to the highest in decades. There existed both supply and demand shocks that led to the shutting down of numerous businesses, and subsequent loss of jobs and life. Working-hour losses in 2020 were approximately four times greater than during the global financial crisis in 2009, with an estimated 220 million people becoming unemployed globally. As said by ILO Director-General Guy Ryde, the world is back to the 2015 level of poverty and development. In fact, this is the same time the UN SDGs were first signed. Thus, in the economic sense, we have traveled back in time in the last two years.

Despite the world opening up, the effects of the pandemic are still widespread. To ensure the livelihood of their citizens, countries need to find ways to combat unemployment. With countries increasing their R&D expenditure, it is necessary to establish that this would lead to a decrease in unemployment in the years to come in order to prevent the unnecessary wastage of money. Out of the vast global expenditure in this realm, about only 10 countries account for 80% of spending. (UNESCO Institute for Statistics, 2021) This raises the question of whether other nations do not have significant proof to commit more to R&D and its benefits to the economy. By finding and establishing the relationship, this paper will help answer the aforementioned question.

The study will utilize cross-sectional data to create simple and multiple linear regression models to draw the relationship between the unemployment rate and R&D expenditure of the country. A negative relation is hypothesized: a greater R&D expenditure per GDP now will lead to a lower unemployment rate in the long run. The rationale propelling this hypothesis is that R&D will not only improve current industry practices like manufacturing but also identify new industries that would require a certain skill set. Hence, more people with the necessary skills will work for these industries, and some would enroll in universities to develop these skills. The second group would be considered out of the labor force and not included when calculating the unemployment rate, thereby lowering the rate. Thus a greater expenditure should boost the number of people hired or in schools, and lower the unemployment rate.

II. Literature Review

To answer the question of whether R&D is good for employment, Piva and Vivarelli (2017) researched R&D expenditures of 674 of the top one thousand R&D investing manufacturing and services firms over the period 2002-2013. The study divided firms into three categories and also independently studied the effects on their number of employees: High-tech when their R&D intensity was larger than 5%, Medium-tech if R&D intensity was between 1% and 5%, and Low-tech if the R&D intensity was less than 1%. The other variables tested were the cost of labor, output, and investment. The researchers then compiled this longitudinal data, and using LSDVC (Least Squares Dummy Variable Corrected) regression technique concluded that there exists a significant labor-friendly impact of R&D expenditure. This effect was in fact found to be limited to only High-tech, and Medium-tech firms with no major changes in Low-tech firms. Further, employment was negatively related to the cost of labor, and positively related to both output and investment for three types of firms. In the end, the study answered what it sought to answer, and proved that R&D expenditure is good for the European Union, at least at the firm level. It also highlighted that this benefit might be skewed only to the firms which invest more in R&D, cementing the hypothesis. However, caution must be taken while generalizing this result. This result only focused on firms and did not give a macroeconomic outlook of aggregating investments by governments and individuals. Further, the firms tested were the top one thousand firms (R&D investment-wise) and were not representative of the economy.

Similar research was conducted by Ciftcioglu and Sokhanvar (2020) on the effects of increasing R&D Intensity to lower the unemployment rate at a country-level in contrast to firm-level research previously conducted. The study analyzed annual data of five European countries from the period 1991-2017 and employed ARDL (Autoregressive Distributed Lag) bounds testing and PMG estimation. The independent variables were percentage growth of GDP (GDPG), inflation rate, and R&D intensity. Inflation was found to be statistically insignificant of a factor in the long run, while GDPG was positively correlated in the long run. R&D investment was also found to decrease the unemployment rate in four out of five countries in the long run. However, in the short run, this investment could further decline employment. The paper argues this is primarily because in the short run R&D investment tries to find ways to reduce output costs by utilizing fewer factors of production, primarily labor. A complementary effect is the creation of new industries and technology which would increase employment. This could lead to a mismatch in the skills of the labor force and the skills required by these new jobs. Thus, this study raises caution against multiple research showing the positive effects of R&D investment. If the R&D investment leads to technological change in the form of automation rather than task creation, the unemployment rate will rise. Another question that this research proposes is how long does it take for R&D investment to affect the employment rate positively. Moreover, should countries invest in R&D extensively to counteract the effects of economic crises like the COVID-19 pandemic? Here again, attention must be given to the sample set considered which is merely five European countries. These results might not generalize well to other countries outside the European Union.

To answer the question raised above about R&D lag, a study was conducted (Brussels 2008) that provided multiple indicators that R&D expenditure lags GDP growth. For private R&D, this lag was around 1 year which is much less compared to the lag of public R&D which was 3-5 years. However, this study had GDP growth as the dependent variable, and not the unemployment rate. This study will also test whether a similar three-year average lag exists for the unemployment rate by comparing R&D investment done three years ago to the collection of the unemployment rate data.

In order to estimate the effect of R&D investment on the unemployment rate, it is crucial to identify other economic factors that affect unemployment so that they can be accounted for while making models, and interpreting results. Abugamea (2018) researched economic factors that influence the unemployment rate. The research used time-series data from 1994-2017 and modeled the relation using multiple linear regression. Three statistically significant factors were found to exist. The GDP growth harmed the unemployment rate, inflation which had a positive impact, and population growth which also had a positive impact. Aurangzeb and Khola (2013) researched the determinants of unemployment in India, China, and Pakistan using OLS analysis, and found GDP growth, population growth, and exchange rate significantly impact the unemployment rate. Since these research were conducted for specific countries, the relationship among the variables might be different on a per-country basis due to different geopolitical and social factors. This research considers this and utilizes many of these factors amongst more (based on data availability) to correctly model the relationship.

There is clearly some research investigating unemployment rates and R&D investment. Some believe the relationship will always be negative, some argue that it depends on the type of technological changes the R&D investment will bring about, and some argue it depends on the time frame we are looking at. Thus, ambiguity exists in the effects of R&D investment. This is exactly what this study aims to clarify. Further, most of the research conducted is based on countries or firms in the European Union. This study by not only focusing on countries in the European Union will answer whether the effects of R&D investment can be seen in other nations of the world. The introduction of a dummy variable will also allow the interpretation of the differences based on developed and developing economies. Additionally, by focusing on R&D investment as a percentage of GDP, a macroeconomic viewpoint will be established, which considers both public and private investment. This will broaden the scope of the results concluded by firm-level research. Lastly, by putting a lag of three years between the two data points, it will see if three years is a significant time for R&D investment to impact unemployment either positively or negatively. Thus, this research should provide a more thorough view of the impact of R&D expenditure on the nation's unemployment rate.

III. Data

To model the relationship between the unemployment rate and expenditure on Research and Development (R&D), cross-sectional data was gathered from 71 countries from across the globe. The dependent variable was chosen to be the number of people unemployed as a percentage of the country's total labor force (*UE*). Labor force refers to the subset of the population who are either working or actively looking for work. The data for unemployment was taken from the World Bank and is for the year 2019. More recent data is not purposefully chosen due to the unprecedented economic challenges faced by the COVID-19 pandemic. The results found in this research suit the more customary atmospheres and do not model economic crises. The main explanatory variable used is the Research and Development Expenditure as a percentage of the country's Gross Domestic Product (*GERD*) expressed in USD. This expenditure includes capital and current expenditures in four sectors: Business enterprise, Government, Higher education, and Private non-profit. R&D covers basic research, applied research, and experimental development. The data was again taken from the World Bank. A larger economy would be able to invest more in R&D compared to a smaller economy despite investing a fewer percentage of their GDP. To account for this, this variable has been chosen instead of total R&D expenditure per country. *GERD* data was collected for 2016 to allow for the three years time lag. The total sample size is 71 countries from North America, South America, Europe, Africa, and Asia. This list can be found in Appendix A. To model the initial relationship, STATA was used to create a scatter plot. This shows a mild negative relationship between *UE* and *GERD*.

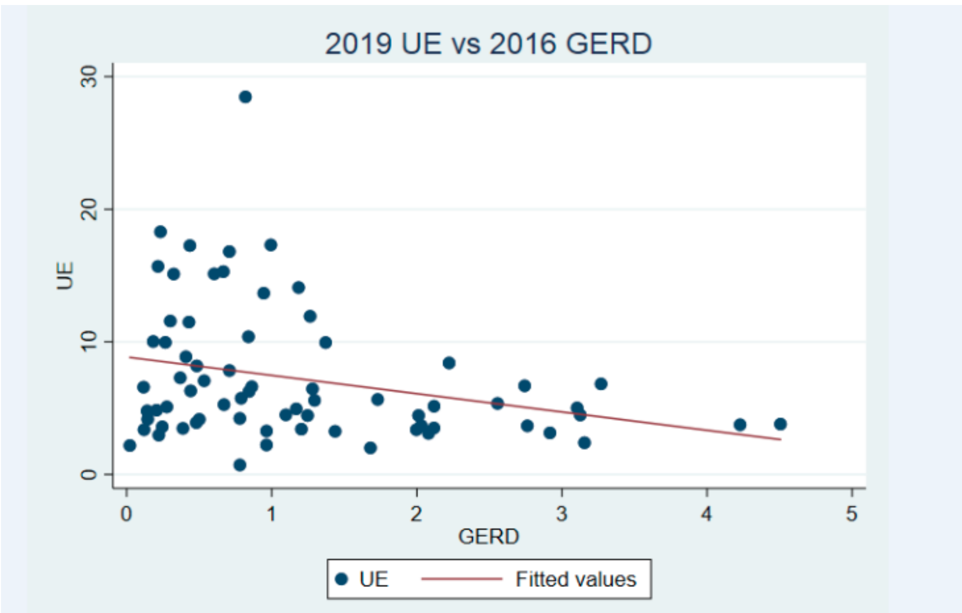

Figure 1 - Scatterplot of UE vs GERD

Apart from the main explanatory variable, *GERD*, the research utilized annual GDP growth rate (*GDPG)*, the logarithm of total population (*LOGPOP),* inflation *(INF),* education index *(EDU)*, compensation of employees as a percent of expense (*COMP),* manufacturing sector's value-added as a percentage of GDP (*MANA*), and service sector's value-added as a percentage of GDP (*SER*) as explanatory variables. Further, a dummy variable for status (*DEV)* was considered to study the statistical differences between developed and developing countries

The *GDPG* data was collected from World Bank's 2019 database, the same year from when our *UE* data was collected. This is done to eliminate job creation effects of a larger economy that are not due to R&D expenditure and maintain the ceteris paribus assumption. In fact, *GDPG* has been empirically proven to affect the unemployment rate. Okun's law states that a 4% increase in GDP will lead to a 1% decrease in the unemployment rate. Hence, it was necessary to include this variable during model formation. Further, population data from World Bank's 2019 database is assessed. This variable is used to eliminate population differences. We take the logarithm of this variable, *LOGPOP*, for easy analysis. A higher *LOGPOP* implies more people wanting to work, which would lead to more humanistic competition, increasing unemployment. The coefficient in the model should hence be positive. Another variable seen in the model is *INF* which is measured by the consumer price index. It is known that a higher inflation rate corresponds to a higher market output. The aggregate demand is more than the aggregate supply, and firms tend to hire more workers to meet this demand. High employment corresponds to lower unemployment. Therefore, a negative coefficient is implied for inflation in our model. This is exactly what Phillip's curve explains. Further, *EDU* is considered which was taken from UNDP (2019). It is the average of mean years of schooling (of adults) and expected years of schooling (of children), both expressed as an index obtained by scaling with the corresponding maxima. *EDU* is used to consider the effects of education level on unemployment. A higher education level means more workers have the necessary skill set to do a job, and consequently, more will be hired, leading to a decrease in the unemployment rate. This should correspond to a negative coefficient in the regression model. Additionally, compensation of employees is also considered (*COMP*) to take into account the effects of salary and minimum wage on employees. This relation is positive, negative, and even zero in multiple research papers. (Manning, 2021) Thus, the coefficient of *COMP* is not hypothesized but is taken to consider its effects on unemployment. Lastly, to consider and eliminate differences between countries based on dominant economic sectors - manufacturing or service - variables *MANA* and *SER* are considered. The coefficients on *MANA* and *SER* are expected to be positive. However, the former is expected to be more in magnitude as the sector can employ more people due to a lower education level requirement.

Below is a summary of the variables that will be utilized in the regression model:

Table 1 - Variable Descriptions

Variable Name	Description	Year	Units	Source
UE	Percentage of Labor Force Unemployed	2019	Percentage	World Bank
GERD	R&D expenditure as a percentage of Gross Domestic Product	2016	Percentage	World Bank
MANA	Manufacturing, value added (% of GDP)	2019	Percentage	World Bank
SER	Services, value added (% of GDP)	2019	Percentage	World Bank
GDPG	Annual GDP Growth Rate	2019	Percentage	World Bank
COMP	Compensation of employees (% of expense)	2019	Percentage	World Bank
LOGPOP	Logarithm of Total Population	2019	Number of People	World Bank
INF	Inflation	2019	Percentage	World Bank
EDU	Education Index (average of mean years of schooling of adults and expected years of schooling of children)	2019	Years	UNDP
DEV (dummy variable)	Development Status of Countries	2019	Dummy: 0 = developing 1 = developed	United Nations

Table 2 - Descriptive Statistics

Variable	Observations	Mean	Std. Dev.	Minimum	Maximum
UE	71	7.23	5.07	0.72	28.47
GERD	71	1.18	1.04	0.02	4.50
MANA	69	13.86	5.61	3.72	32.01
SER	70	59.94	7.91	36.97	79.16
GDPG	71	2.77	1.80	-0.40	9.46
COMP	61	18.99	9.85	5.82	49.66
LOGPOP	71	16.37	1.78	11.49	21.06
INF	71	2.61	2.57	-1.93	15.17
EDU	71	0.78	0.11	0.46	0.94
DEV	71	0.39	0.49	0	1

Here, the six Classical Linear Model (CLM) assumptions are first tested before using linear regression models:

1. **Model is linear in its parameters:** : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$
   $\beta_{0,}$ β_1, β_2,.., and β_k are the slope parameters that quantify the relationship between the dependent variable and the explanatory variables. The u represents the error/ disturbance term and is the intercept of the linear regression line. The degree of this polynomial is 1 which means it is linear, so the assumption is valid.
2. **Random Sampling of data:**
   The 71 countries were chosen from the set of all countries that had all the data points. This was done randomly without any bias to a continent. The sample has multiple countries from the same continent, both smaller and larger economies, and developed and developing nations. This is random and satisfies the assumption.
3. **No perfect collinearity:**
   This was tested using the STATA correlation table which can be found in Appendix B. Clearly none of the variables are perfectly correlated with correlation values less than one. Further, none of the variables were constant, each having a standard deviation not equal to zero. This assumption is therefore satisfied.
4. **Zero Conditional Mean**
   The zero conditional mean assumption means that the expected value of u, the error term, should be equal to zero given any of the independent variables. That is $E(u_i x_i) = 0$, for all i =

1, 2,..., n. As seen multiple factors are affecting our dependent variable, UE. The research has tried to identify most of them and assessed them during the multiple linear regression model. After calculating the residuals, the mean comes out to be $-9.19 \cdot e^{-9}$, which is very close to zero. Thus, this condition can be assumed to be satisfied by the dataset.

5. **Homoskedasticity**

   This means that the variance in the error term, u should be a constant value given any of the independent variables. To check for homoskedasticity, a graph between the residuals and fitted values is created (Figure - 2). It can be seen that the spread of the residuals for the most part is constant (equidistant above and below the zero line). However, as the fitted values grow bigger, the variance can be seen increasing due to outliers. As such, the results are interpreted with caution.

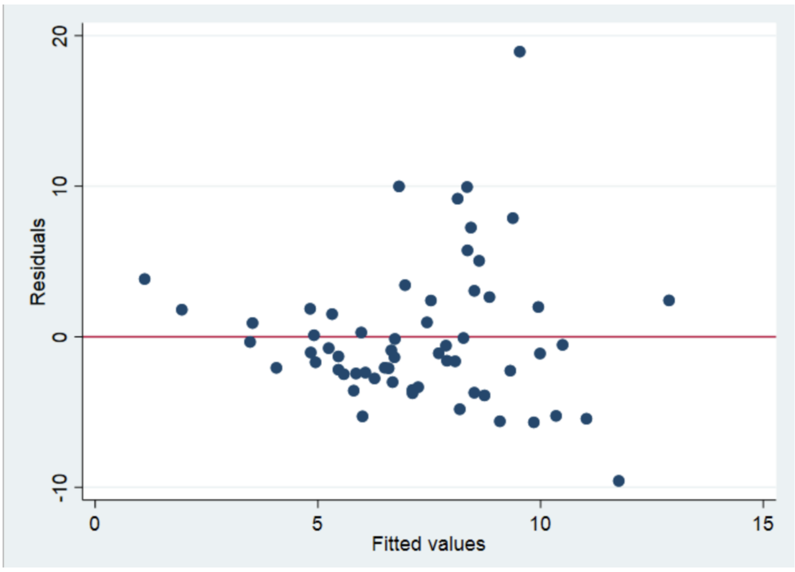

Figure 2 - Scatterplot of residuals vs fitted values

6. **Normality**

   This means that the population error u should be independent of the explanatory variables and be normally distributed with zero mean and variance σ^2: $u \sim \text{Normal}(0, \sigma^2)$. This assumption is tested by plotting a graph of the residuals in Figure 3. The residuals seem to approximately follow a normal distribution. Further, Figure 4 shows a QQ plot. This graph is used to compare two probability distributions. The residuals are seen to slightly diverge from a true normal distribution and the ends (the outliers). Hence, this assumption may not be completely valid for this data set and results are interpreted with caution.

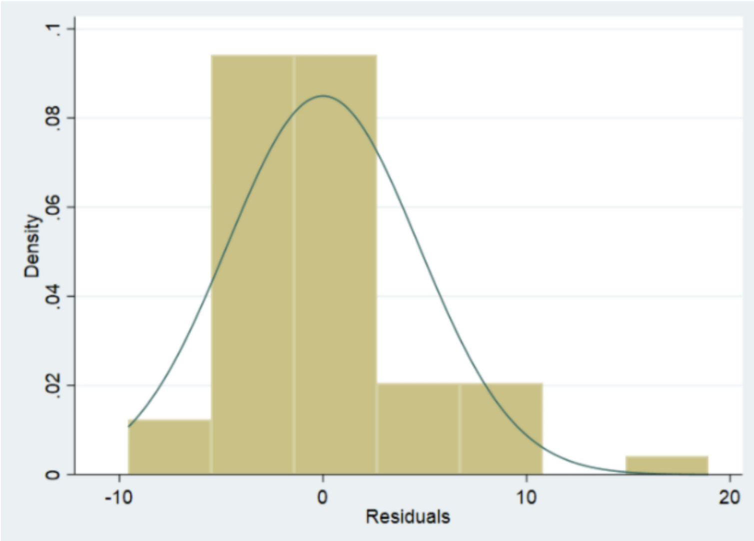

Figure 3 - Histogram of residuals

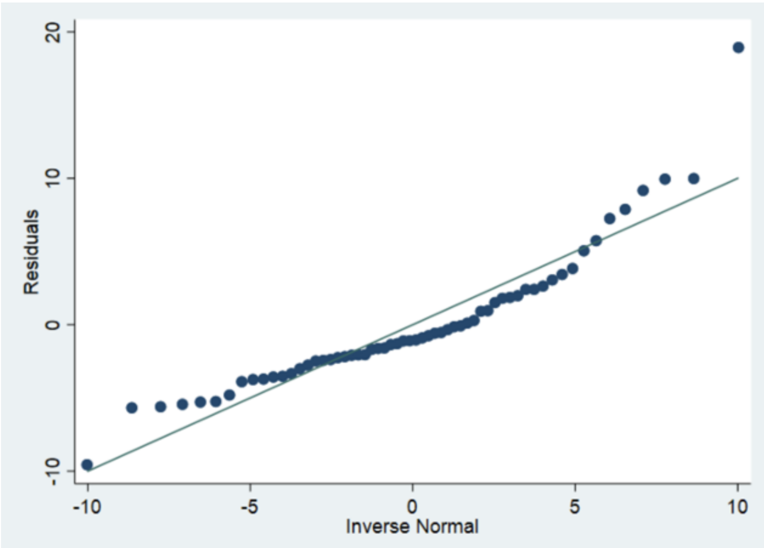

Figure 4 - QQ plot

IV. Results

After seeing how all the six Classical Linear Model (CLM) assumptions hold for the data set, we can begin formulating different linear regression models for analysis.

<u>Model-1:</u>

This model is the simple linear regression model which considers the effect of our main independent variable, GERD, and the dependent variable, UE. All other factors are taken as part of the standard error term, u. This model can be written as:

$$UE = \beta_0 + \beta_1(GERD) + u$$

The data was collected for 71 countries. Using STATA, the regression coefficients identified are:

$$UE = 8.856 - 1.381(GERD)$$
$$n = 71,\ R^2 = 0.08$$

The slope parameter is highly negative as initially hypothesized. This level-level model indicates that an increase in the R&D expenditure as a percentage of GDP by 1 unit, will lead to a 1.381 decrease in the unemployment rate as a percentage of the total labor force. Using the p-value of the variable (0.017), *GERD* comes out to be statistically significant at the 1.7% level which is encouraging. The R-squared value of the model is 0.08 which means that a less proportion of variation in *UE* can be explained by solely considering *GERD*. More factors need to be considered, and other explanatory variables need to be controlled to accurately estimate the impact of *GERD* on *UE*. This is exactly done by the multiple linear regression model.

<u>Model-2:</u>
This model considers all other secondary explanatory variables: *GDPG, MANA, LOGPOP, INF, EDU, SER,* and *COMP.* The equation is:

$$UE = \beta_0 + \beta_1(GERD) + \beta_2(GDPG) + \beta_3(MANA) + \beta_4(LOGPOP) + \beta_5(INF) + \beta_6(EDU) + \beta_6(SER) + \beta_6(COMP) + u$$

The data input here is from 60 countries due to the lack of data for some countries. STATA's estimated coefficients for this equation is:

$$UE = 22.098 - 0.658(GERD) - 0.047(GDPG) - 0.170(MANA) - 0.226(LOGPOP) + 0.147(INF) - 16.380(EDU) + 0.103(SER) - 0.072(COMP)$$
$$n = 60,\ R^2 = 0.18$$

The model has an R-squared value of 0.18, which is higher compared to 0.08 of the previous model. The explanatory variables utilized can explain 18% of the variation in the independent variable. This model has a coefficient for *GERD* as -0.658 compared to -1.381 earlier. This is a lesser value, but still negative as hypothesized. This level-level model indicates that an increase in the R&D expenditure as a percentage of GDP by 1 unit, will lead to a 0.658 decrease in the unemployment rate as a percentage of the total labor force keeping everything else constant.

As for the other explanatory variables, the coefficients are both positive and negative. For *GDPG, LOGPOP*, *MANA*, *INF, EDU,* and *COMP* the coefficients are negative while for *INF* and *SER* the coefficients are positive. What is interesting here is the coefficient of *LOGPOP, INF,* and *SER.* All the other variables influence *UE* in the same direction as initially hypothesized in the previous section. The coefficient of *LOGPOP* however is -0.170 meaning a 1 percent increase in the total population will lead to a decrease in the unemployment rate by 0.226 percent keeping other variables constant. A reason for this negative relationship could be that with a larger population, more jobs are created like home building, teaching, etc. This might overcome the competition created and lead to lower unemployment. Further, the coefficient of *INF* was 0.147, meaning a 1 point increase in inflation will increase the unemployment rate by 0.147 percent. This is contrary to the Phillips curve, and a reason might be the tighter monetary control by governments to make sure inflation does not rise to high levels. This result supports modern research on the flattening of the Phillips curve. (Kuttner & Robinson, 2010)

Based on the p values, none of the explanatory variables came out to be significant at the 10% level. This could be because of lack of data from more countries and possible multicollinarity between *GERD* and *EDU*. This will be further explored in the extension section of this paper. However, looking at the initial scatterplot (Figure - 1) there seems to be a significant negative relation between *GERD* and *UE*. Thus, a new model is formulated as follows.

Model-3:

This model considers *GERD* as the main explanatory variable, with *MANA, INF,* and *EDU* as secondary explanatory variables. The manufacturing sector's value-added was chosen in this model to account for differences between the economies of the country. There is documented evidence of the large ripple effects caused by the loss of manufacturing sector jobs in other sectors. Bivens (2019) calculated that for every 100 jobs lost in durable manufacturing, there are 744.1 indirect jobs lost. Thus, this variable is chosen. Further, *INF* is chosen due to multiple research pointing to its negative effects on unemployment (Phillip's curves). Lastly, *EDU* was taken into consideration as it was the most significant variable in Model-2 with a p-value of 0.11, meaning significance at the 11% level. The model is:

$$UE = \beta_0 + \beta_1(GERD) + \beta_2(MANA) + \beta_3(INF) + \beta_4(EDU) + u$$

On running through STATA, the regression model parameters come out to be:

$$UE = 16.984 - 0.458(GERD) - 0.231(MANA) + 0.203(INF) - 8.287(EDU)$$
$$n = 69,\ R^2 = 0.16$$

The model has an R-squared of 0.16 meaning the explanatory variables utilized can explain 16% of the variation in the independent variable. This is lower compared to 0.18 of the previous model which was expected as the squared sum of residuals (SSR) decreases on adding more independent variables, which increases R-squared. (R^2 = 1 - SSR/SST) This model has a coefficient for *GERD* as -0.458 compared to -.658 earlier. This is a lesser value, but still negative as hypothesized. This level-level model indicates that an increase in the R&D expenditure as a percentage of GDP by 1 unit, will lead to a 0.458 decrease in the unemployment rate as a percentage of the total labor force keeping everything else constant. Further, the coefficients of *MANA*, and *EDU* were negative as hypothesized. However, here again, the coefficient of *INF* is positive, contradicting the Phillips Curves. This again provides evidence in support of the flattening of the curve.

Looking at the p values of the explanatory variables, we only find *MANA* to be statistically significant at the 3.9 % level. This highlights how a larger manufacturing sector corresponds to a lower unemployment rate. Moving forward, this variable will be considered in the models.

<u>Model-4:</u>

In this model, all the statistically insignificant variables of the previous model except for the primary explanatory variable are dropped. Thus, only *GERD* and *MANA* are considered. The model is:

$$UE = \beta_0 + \beta_1(GERD) + \beta_2(MANA) + u$$

On running through STATA, the regression model comes out to be:

$$UE = 11.428 - 1.132(GERD) - 0.202(MANA)$$
$$n = 69,\ R^2 = 0.13$$

This model has an R-squared of 0.13, meaning the explanatory variables utilized can explain 13% of the variation in the independent variable. This will be less than the previous model as we dropped two explanatory variables as they were insignificant. The coefficient of *GERD* increases to -1.132 which implies that an increase in the R&D expenditure as a percentage of GDP by 1 unit, will lead to a 1.132 decrease in the unemployment rate as a percentage of the total labor force keeping everything else constant. Further, an increase in the manufacturing sector's value-added by 1 unit, will lead to a 0.202 point decrease in the unemployment rate as a percentage of the total labor force ceteris paribus.

Looking for statistical significance, both the variables come out to be significant with p - values 0.05 and 0.06 for *GERD* and *MANA.* This means *GERD* is significant at the 5% level while *MANA* is significant at the 6% level.

The results summary for the four models is below:

Results summary

Dependent Variable *UE*				
Independent Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**
GERD	-1.38** (0.56)	-0.66 (0.90)	-0.46 (0.73)	-1.13** (0.58)
MANA	--	-0.17 (0.15)	-0.23** (0.11)	-0.20** (0.11)
GDPG	--	-0.05 (0.43)		
COMP	--	-0.07 (0.10)		
LOGPOP	--	-0.23 (0.59)		
INF	--	0.15 (0.30)	0.20 (0.24)	
EDU	--	-16.38 (10.1)	-8.29 (6.76)	
SER	--	0.10 (0.10)		
Intercept	8.86*** (0.88)	22.10 (16.83)	16.98*** (5.34)	11.43*** (1.60)
Observations	71	60	69	69
R-squared	0.08	0.19	0.16	0.13

*Significant at 10%, ** 5%, *** 1%

V. Extensions

In Model - 3, it is evident that only *MANA* is statistically significant, while *GERD* and *EDU* are not. However, on the removal of *EDU* from the model, both *GERD* and *MANA* come out significant as shown by Model - 4. Further, on the removal of *GERD,* both *EDU* and *MANA* are significant. (Model in Appendix - C) On looking through the correlation table (Appendix - B), *GERD* and *EDU* are moderately correlated (correlation coefficient of 0.61). Thus, to test robustness we perform an F-test with the hypothesis:

$$H_0 : \beta_1 = \beta_3 = 0$$
$$H_1 : H_0 \text{ is false}$$

The unrestricted model utilized is:

$$UE = \beta_0 + \beta_1(GERD) + \beta_2(MANA) + \beta_3(EDU) + u$$
$$n = 69, R^2 = 0.15$$

The restricted model utilized is:

$$UE = \beta_0 + \beta_2(MANA) + u$$
$$n = 69, R^2 = 0.08$$

F - Test:

$$F = \frac{[(R_{ur}^2 - R_r^2)/q]}{[(1-R_{ur}^2)/(n-k-1)]}$$

Using the F-test based on the above diagram, the F statistic is 2.68, which is greater than the critical value of $F_{2, 65} = 2.63$ at the 8% level. This implies that *GERD* and *EDU* are jointly significant at the 8% level, and due to multicollinearity, both turn out to be individually insignificant in the regression models utilizing both variables. A direct conclusion of this result is that both *GERD* and *EDU* are important factors that affect the unemployment rate of a country. Model - 4 removes *EDU* to deal with multicollinearity. In future studies, both variables should be considered together in an appropriate model.

Another extension to Model - 4 is the addition of a dummy variable, *DEV* which is a way of classifying countries based on the level of development as proposed by the World Economic Situation and Prospects report (2014). In the model, *DEV* has a value of 0 for economies in transition and developing economies, and 1 for developed economies. The dummy variable is added to the last model, forming Model - 5 as follows:

$$UE = \beta_0 + \beta_1(GERD) + \beta_2(MANA) + \beta_3(DEV) + u$$

After running through STATA, the regression model comes out to be:

$$UE = 12.075 - 0.705(GERD) - 0.235(MANA) - 1.777(DEV)$$
$$n = 69,\ R^2 = 0.15$$

Here, the R - squared value increases to 0.15 which is greater than that of Model - 4. The coefficients of *GERD* and *MANA* still remain negative, while the coefficient for *DEV* was also negative. This points to the fact that developed economies will have unemployment rates less by 1.777 points compared to developing economies. During the statistical significance test for the variables, *GERD* and *DEV* both turned out to be insignificant at the 10% level with p values 0.292 and 0.204. However, *MANA* became more significant with a p-value of 0.038. This shows that there are no statistically significant differences between developed and developing countries based on the percentage of GDP used for R&D expenditure. Other factors like the size of the manufacturing sector cause the found variation in the dependent variable.

VI. Conclusions

The research study was conducted with the hypothesis that R&D expenditure as a percentage of GDP will negatively affect the unemployment rate of countries. From Model-1 to Model-5, it can be seen that the coefficient of *GERD* has always been negative, proving the initial hypothesis. Model-4 regressed the unemployment rate on R&D expenditure and manufacturing sector's service added and found both to be significant ~5% level. A one-point increase in the R&D expenditure as a percentage of GDP was found to decrease the unemployment rate by 1.132 ceteris paribus. All other variables were found to be statistically insignificant in the various models.

Further, the data for *GERD* was taken from 2016 while *UE* was taken from 2019. As the effects became noticeable after this three-year lag, it can be concluded that this is an appropriate time horizon that countries can consider while deciding to invest in R&D. Moreover this research highlights that countries should explore other channels to reduce unemployment in the short run caused by crises like the COVID-19 pandemic.

Another interesting result of this research is the high statistical significance of the manufacturing sector's service added, *MANA,* on the unemployment rate in Model - 4. A one-point increase in *MANA* was found to decrease *UE* by 0.2 points ceteris paribus. This implied that countries with a large manufacturing sector tend to have lower unemployment. So, if the "task-creation" effects of R&D outweigh the "automation" effects, *UE* will decrease due to the double effect of both high *GERD* and larger *MANA.*

As further extensions to the model, the joint significance of the country's education level, *EDU,* and *GERD* was considered in an F-test. They both were found to be jointly significant at the 8% level. Due to their moderate collinearity, their individual statistical significance was mutually exclusive in multiple models. The f-test reinforced their significance and showed both *GERD* and *EDU* effect *UE.* To add another dimension to the study, model - 5 was also created with a dummy variable *DEV* as a proxy for development status. *GERD* came out to be statistically insignificant while considering this variable showing that there is no major difference in unemployment rates between these countries due to differences in R&D expenditure.

This study shines a light on the complexity of systems that affect the unemployment rate and how it is hard to single out the effect of one variable. This can be shown by the high statistical significance of the intercept term in the final regression model. As the data was considered on a per-country basis, a low R-squared value was predicted and found (Model - 4, R^2=0.13) This is primarily due to the differences in the economical landscape between countries and that only 69 countries were taken into consideration. To better represent the relation, future studies should analyze data from a larger set of countries, and efforts should be made to identify more independent variables that influence *UE* and have low correlations to *GERD*. Nonetheless, this research can represent a base for further research as it shows that R&D expenditure has a negative relation with the unemployment rate. This makes it a possible instrument that countries can utilize to fight unemployment in the long run.

References:

1) Abugamea, D. G. (2018). Determinants of Unemployment: Empirical Evidence from Palestine. *MPRA Paper,* 89424, University Library of Munich, Germany.
2) United Nations. (n.d.). *Goal 9 | Department of Economic and Social Affairs*. United Nations. Retrieved October 19, 2021, from https://sdgs.un.org/goals/goal9.
3) United Nations. (2021, June 2). *Covid crisis to push global unemployment over 200 million mark in 2022 | | UN news*. United Nations. Retrieved October 19, 2021, from https://news.un.org/en/story/2021/06/1093182.
4) UNESCO Institute for Statistics. (2021). *How much does your country invest in R&D* Retrieved October 19, 2021, from http://uis.unesco.org/apps/visualisations/research-and-development-spending/.
5) Piva, M., & Marco, V. (2017). Is R&D Good for Employment? Microeconometric Evidence from the EU*, IZA Discussion Papers, No. 10581,* Institute of Labor Economics (IZA), Bonn. Retrieved October 20, 2021, from https://www.econstor.eu/bitstream/10419/161204/1/dp10581.pdf
6) Serhan Ciftcioglu, S. C., & Sokhanvar, A. S. (2020). Can Increasing the R&D Intensity Lower Unemployment Rate? Case of Five Selected European Countries. *Ekonomický Časopis*, 68(2), 188–207. https://www.researchgate.net/publication/340923438_Can_Increasing_the_RD_Intensity_Lower_Unemployment_Rate_Case_of_Five_Selected_European_Countries
7) A Time Series Analysis of the Development in National R&D Intensities and National Public Expenditures on R&D. (2018). *Rindicate*. Retrieved from https://ec.europa.eu/invest-in-research/pdf/download_en/spa4_final_report_final.pdf.
8) Riddell, W. C. (1979). The empirical foundations of the Phillips curve: Evidence from Canadian Wage Contract Data. *Econometrica*, *47*(1), 1. https://doi.org/10.2307/1912343
9) World Economic Situation and Prospects. (2014). *Country classification - United Nations.* Retrieved November 23, 2021, from https://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf.
10) Bivens , J. (2019, January 23). *Updated employment multipliers for the U.S. economy*. Economic Policy Institute. Retrieved November 15, 2021, from https://www.epi.org/publication/updated-employment-multipliers-for-the-u-s-economy/.
11) Manning, A. (2021). The elusive employment effect of the minimum wage. *Journal of Economic Perspectives*, *35*(1), 3–26. https://doi.org/10.1257/jep.35.1.3

12) Kuttner, K., & Robinson, T. (2010). Understanding the flattening Phillips curve. *The North American Journal of Economics and Finance*, *21*(2), 110–125. https://doi.org/10.1016/j.najef.2008.10.003

Appendix A: List of 71 countries used in the creation of the models

Armenia	Chile	El Salvador	Iceland	Korea, Rep.	Netherlands	Russian Federation	Sweden
Austria	China	Estonia	India	Latvia	North Macedonia	Rwanda	Thailand
Azerbaijan	Colombia	Finland	Indonesia	Lithuania	Norway	Serbia	Tunisia
Belarus	Costa Rica	France	Ireland	Luxembourg	Panama	Seychelles	Turkey
Belgium	Croatia	Georgia	Israel	Malaysia	Paraguay	Singapore	Ukraine
Bosnia and Herzegovina	Cyprus	Germany	Italy	Mexico	Peru	Slovak Republic	United Arab Emirates
Brazil	Czech Republic	Greece	Japan	Moldova	Poland	Slovenia	United States
Bulgaria	Denmark	Guatemala	Jordan	Mongolia	Portugal	South Africa	Uruguay
Canada	Egypt, Arab Rep.	Hungary	Kazakhstan	Montenegro	Romania	Spain	

Appendix B: Correlation matrix used to show how Gauss-Markov assumption 3 is satisfied

	UE	*GERD*	*SER*	*GDPG*	*COMP*	*LOGPOP*	*INF*	*EDU*	*MANA*	*DEV*
UE	1.00									
GERD	-0.27	1.00								
SER	-0.17	0.43	1.00							
GDPG	0.05	-0.21	-0.29	1.00						
COMP	0.05	-0.35	-0.04	-0.001	1.00					
LOGPOP	-0.01	0.07	-0.12	-0.21	-0.29	1.00				
INF	0.12	-0.29	-0.22	-0.05	-0.13	0.20	1.00			
EDU	-0.30	0.61	0.38	-0.28	-0.45	-0.20	-0.14	1.00		
MANA	-0.28	0.13	-0.27	0.02	0.10	0.23	-0.01	0.05	1.00	
DEV	-0.20	0.43	0.52	-0.16	-0.33	-0.21	-0.28	0.61	-0.17	1.00

Appendix C: STATA Regression Model Outputs

Model - 1:

```
. regress ue rd

      Source |       SS           df       MS      Number of obs   =        71
-------------+----------------------------------   F(1, 69)        =      6.02
       Model |  144.383143         1  144.383143   Prob > F        =    0.0167
    Residual |  1655.71429        69  23.9958592   R-squared       =    0.0802
-------------+----------------------------------   Adj R-squared   =    0.0669
       Total |  1800.09743        70  25.7156776   Root MSE        =    4.8986

------------------------------------------------------------------------------
          ue |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          rd |  -1.380756   .5628942    -2.45   0.017    -2.503699   -.2578126
       _cons |   8.855803   .8808922    10.05   0.000     7.098471    10.61313
------------------------------------------------------------------------------
```

Model - 2:

```
. regress ue rd gdpg mana logpop inf edu ser comp

      Source |       SS           df       MS      Number of obs   =        60
-------------+----------------------------------   F(8, 51)        =      1.47
       Model |  299.042679         8  37.3803349   Prob > F        =    0.1932
    Residual |  1300.84903        51  25.5068438   R-squared       =    0.1869
-------------+----------------------------------   Adj R-squared   =    0.0594
       Total |  1599.89171        59  27.1168087   Root MSE        =    5.0504

------------------------------------------------------------------------------
          ue |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          rd |  -.6583462   .9041112    -0.73   0.470    -2.473425    1.156733
        gdpg |  -.0470209   .4261054    -0.11   0.913    -.9024632    .8084213
        mana |  -.1703088   .1470539    -1.16   0.252    -.4655318    .1249143
      logpop |  -.2257952   .5938208    -0.38   0.705     -1.41794    .9663497
         inf |   .1465483   .2953901     0.50   0.622    -.4464721    .7395687
         edu |  -16.38001   10.10049    -1.62   0.111     -36.6576    3.897581
         ser |   .1026218   .1098788     0.93   0.355     -.117969    .3232126
        comp |  -.0723122   .0951735    -0.76   0.451    -.2633809    .1187566
       _cons |   22.09819    16.8336     1.31   0.195    -11.69668    55.89306
------------------------------------------------------------------------------
```

Model - 3:

```
. regress ue inf rd edu mana

      Source |       SS           df       MS      Number of obs   =        69
-------------+----------------------------------   F(4, 64)        =      3.02
       Model |  284.159479         4  71.0398697   Prob > F        =    0.0240
    Residual |  1504.14784        64    23.50231   R-squared       =    0.1589
-------------+----------------------------------   Adj R-squared   =    0.1063
       Total |  1788.30732        68   26.298637   Root MSE        =    4.8479

------------------------------------------------------------------------------
          ue |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inf |   .2034746   .2373233     0.86   0.394    -.2706331    .6775824
          rd |  -.4578747   .7310607    -0.63   0.533    -1.918336    1.002587
         edu |  -8.287252   6.763925    -1.23   0.225    -21.79974    5.225241
        mana |  -.2305308    .109611    -2.10   0.039    -.4495039   -.0115576
       _cons |   16.98356   5.339811     3.18   0.002     6.316067    27.65106
------------------------------------------------------------------------------
```

Model - 4:

```
. regress ue rd mana

      Source |       SS           df       MS      Number of obs   =        69
-------------+----------------------------------   F(2, 66)        =      4.83
       Model |  228.340303         2  114.170151   Prob > F        =    0.0110
    Residual |  1559.96702        66  23.6358639   R-squared       =    0.1277
-------------+----------------------------------   Adj R-squared   =    0.1013
       Total |  1788.30732        68   26.298637   Root MSE        =    4.8617

------------------------------------------------------------------------------
          ue |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          rd |  -1.132144   .5776107    -1.96   0.054    -2.285381    .0210927
        mana |  -.2020828   .1083521    -1.87   0.067    -.4184147     .014249
       _cons |   11.42766   1.600962     7.14   0.000     8.231241    14.62409
------------------------------------------------------------------------------
```

Model - 5:

```
. regress ue rd mana dev

      Source |       SS           df       MS      Number of obs   =        69
-------------+----------------------------------   F(3, 65)        =      3.80
       Model |  266.944513         3  88.9815045   Prob > F        =    0.0142
    Residual |  1521.36281        65  23.4055816   R-squared       =    0.1493
-------------+----------------------------------   Adj R-squared   =    0.1100
       Total |  1788.30732        68   26.298637   Root MSE        =    4.8379

------------------------------------------------------------------------------
          ue |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          rd |   -.705215   .6639973    -1.06   0.292    -2.031309    .6208791
        mana |  -.2347389   .1107806    -2.12   0.038    -.4559832   -.0134947
         dev |  -1.776916   1.383596    -1.28   0.204    -4.540148    .9863159
       _cons |   12.07455   1.670874     7.23   0.000     8.737587    15.41152
------------------------------------------------------------------------------
```

Models for F-tests :

```
. regress ue mana

      Source |       SS           df       MS      Number of obs   =        69
-------------+----------------------------------   F(1, 67)        =      5.58
       Model |   137.53639         1   137.53639   Prob > F        =    0.0211
    Residual |  1650.77093        67  24.6383721   R-squared       =    0.0769
-------------+----------------------------------   Adj R-squared   =    0.0631
       Total |  1788.30732        68   26.298637   Root MSE        =    4.9637

------------------------------------------------------------------------------
          ue |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        mana |  -.2535741   .1073253    -2.36   0.021    -.4677962   -.0393519
       _cons |   10.81273   1.602868     6.75   0.000     7.613392    14.01207
------------------------------------------------------------------------------
```

```
. regress ue mana edu

      Source |       SS           df       MS      Number of obs   =        69
-------------+----------------------------------   F(2, 66)        =      5.38
       Model |  250.804272         2  125.402136   Prob > F        =    0.0068
    Residual |  1537.50305        66  23.2955007   R-squared       =    0.1402
-------------+----------------------------------   Adj R-squared   =    0.1142
       Total |  1788.30732        68   26.298637   Root MSE        =    4.8265

------------------------------------------------------------------------------
          ue |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        mana |  -.2467112   .1044059    -2.36   0.021    -.4551643    -.038258
         edu |  -11.94888   5.418883    -2.21   0.031    -22.76803   -1.129729
       _cons |   20.06719      4.477     4.48   0.000     11.12857    29.00581
------------------------------------------------------------------------------
```

```
. regress ue rd edu mana

      Source |       SS           df       MS      Number of obs   =        69
-------------+----------------------------------   F(3, 65)        =      3.80
       Model |  266.883189         3  88.9610631   Prob > F        =    0.0142
    Residual |  1521.42413        65  23.4065251   R-squared       =    0.1492
-------------+----------------------------------   Adj R-squared   =    0.1100
       Total |  1788.30732        68   26.298637   Root MSE        =     4.838

------------------------------------------------------------------------------
          ue |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          rd |  -.5909052    .712948    -0.83   0.410    -2.014761    .8329504
         edu |  -8.645413   6.737241    -1.28   0.204    -22.10061    4.809787
        mana |  -.2217335   .1089071    -2.04   0.046     -.439236    -.004231
       _cons |    17.8296   5.237139     3.40   0.001     7.370307    28.28888
------------------------------------------------------------------------------
```